COMPLEX WORD IDENTIFICATION SHARED TASK 2018

*Seid M. Yimam, Chris Biemann, Shervin Malmasi, Gustavo Paetzold, Lucia Specia, Sanja Štajner, **Anaïs Tack**, Marcos Zampieri*

Universität Hamburg
University of Mannheim
The University of Sheffield
Harvard Medical School
University of Wolverhampton
UCL Université catholique de Louvain
KU Leuven

June 5 - BEA13 at NAACL 2018 - New Orleans, Louisiana

COMPLEX WORD IDENTIFICATION (CWI)

automatically identify which of the running words in a text will be found difficult for a target reader

- why?
  - assess global text readability
    (Dale & Chall, 1948; Gunning, 1952; Mc Laughlin, 1969)
  - automatic text simplification
    (Shardlow, 2013, 2014; Paetzold & Specia, 2016a)
- machine learning approach
  - features of word complexity: length, frequency, familiarity, …
    (Elhadad, 2006; Shardlow, 2013; Kauchak, 2016; Wróbel, 2016)
  - gold-standard annotations of complex words
    (Paetzold & Specia, 2016b; Yimam, Štajner, Riedl, & Biemann, 2017a, 2017b)

CWI SHARED TASK AT SEMEVAL 2016 (PAETZOLD & SPECIA, 2016B)

- unknown words for non-natives in English
  - 400 annotators
  - broad scope of complexity to predict individual difficulty
  - complex if at least 1 non-native deemed it complex
- submitted systems
  - ensemble methods/random forests: top-performing (Malmasi, Dras, & Zampieri, 2016; Ronzano, Abura'ed, Espinosa Anke, & Saggion, 2016)
  - neural networks: performance lower than traditional ML
- challenging task when it comes to data annotation (Zampieri, Malmasi, Paetzold, & Specia, 2017)

OUTLINE

1. Complex Word Identification
2. Shared Task
    - Dataset
    - Description
3. Systems
4. Results
    - Monolingual Tracks
    - Multilingual Track
5. Conclusion

SHARED TASK

DATASET

- **CWIG3G2**: three genres, two user groups (Yimam et al., 2017)
- Multilingual dataset
  - **English:** Wikipedia, News, Wikinews
  - **German:** Wikipedia
  - **Spanish:** Wikipedia
  - **French:** Wikipedia (Brouwers et al., 2014)

	train	dev	test
English	27,299	3,328	4,252
German	6,151	795	959
Spanish	13,750	1,622	2,233
French	-	-	2,251

ANNOTATION

- **Amazon Mechanical Turk**
  - 100 HITs (Human Intelligence Tasks) per language
  - 5 to 10 sentences in a single HIT
  - single or multiword annotations

- **native** (NS) and **non-native** (NNS) annotators
  - 20 for English (10 NS & 10 NNS)
  - 10 for German, Spanish and French (mixed NS & NNS)

TASK DESCRIPTION

sites.google.com/view/cwisharedtask2018

➤ 4 tracks

- 3 **monolingual** tracks: English - German - Spanish
- 1 **multilingual** track: EN/DE/ES → French

➤ 2 classification tasks

Both	China	and the	Philippines	flexed	their	muscles	on	Wednesday
	simple		simple	complex	complex	simple		simple
	0.0		0.0	0.4	0.25	0.0		0.0

- **binary**: complex if at least one subject **(F_1 macro)**
- **probabilistic**: ratio of complex annotations **(MAE)**

SYSTEMS

CLASSIFIERS & REGRESSORS FOR COMPLEX WORD IDENTIFICATION

- 12 teams
- 252 runs in total
- variety of methods

types	#	papers
Random Trees / Random Forests / Extra Trees / AdaBoost / XGBoost	8	(Alfter and Pilan, 2018; Kajiwara and Komachi, 2018; Bingel and Bjerva, 2018; AbuRa'ed and Saggion, 2018; Wani et al., 2018; Aroyehun et al., 2018; Gooding and Kochmar, 2018; Hartmann and dos Santos, 2018)
FFNN / LSTM / CNN / embeddings	5	(Hartmann and dos Santos, 2018; De Hertog and Tack, 2018; Bingel and Bjerva, 2018; Wani et al., 2018; Aroyehun et al., 2018)
LibSVM / v-SVR	3	(Butnaru and Ionescu, 2018; AbuRa'ed and Saggion, 2018; Wani et al., 2018)
Voting	2	(Hartmann and dos Santos, 2018; Wani et al., 2018)
Naive Bayes	1	(Popovic, 2018)
Linear Regression	1	(Gooding and Kochmar, 2018)
J48 / REP / LMT / JRip / PART	1	(Wani et al., 2018)

FEATURES FOR COMPLEX WORD IDENTIFICATION

types	#	papers
word length	10	(Butnaru and Ionescu, 2018; Alfter and Pilan, 2018; Hartmann and dos Santos, 2018; Kajiwara and Komachi, 2018; De Hertog and Tack, 2018; Gooding and Kochmar, 2018; Bingel and Bjerva, 2018; AbuRa'ed and Saggion, 2018; Wani et al., 2018)
semantics / WordNet	7	(Butnaru and Ionescu, 2018; Alfter and Pilan, 2018; Hartmann and dos Santos, 2018; Gooding and Kochmar, 2018; Bingel and Bjerva, 2018; AbuRa'ed and Saggion, 2018; Wani et al., 2018)
word frequency / likelihood	7	(Alfter and Pilan, 2018; Kajiwara and Komachi, 2018; De Hertog and Tack, 2018; Bingel and Bjerva, 2018; AbuRa'ed and Saggion, 2018; Wani et al., 2018; Aroyehun et al., 2018)
psycholinguistic	5	(Alfter and Pilan, 2018; Hartmann and dos Santos, 2018; De Hertog and Tack, 2018; Gooding and Kochmar, 2018; Aroyehun et al., 2018)
word embeddings	5	(Butnaru and Ionescu, 2018; Alfter and Pilan, 2018; Hartmann and dos Santos, 2018; De Hertog and Tack, 2018; Bingel and Bjerva, 2018)
n-grams	4	(Alfter and Pilan, 2018; Popovic, 2018; Hartmann and dos Santos, 2018; Gooding and Kochmar, 2018)
syllable count	4	(Alfter and Pilan, 2018; Hartmann and dos Santos, 2018; Gooding and Kochmar, 2018; Wani et al., 2018)
part of speech	3	(Alfter and Pilan, 2018; Gooding and Kochmar, 2018; Bingel and Bjerva, 2018)
dependency relations	2	(Gooding and Kochmar, 2018; AbuRa'ed and Saggion, 2018)
sentence length	2	(AbuRa'ed and Saggion, 2018; Wani et al., 2018)
stemming	2	(Alfter and Pilan, 2018; Bingel and Bjerva, 2018)
vowel count	2	(Butnaru and Ionescu, 2018; Wani et al., 2018)

RESULTS

Monolingual Tracks

ENGLISH – BINARY TASK

Both	China	and	the	Philippines	flexed	their	muscles	on	Wednesday
	simple			simple	complex	complex	simple		simple

	team	*paper*	avg rank	News	Wikinews	Wikipedia
1	**CAMB**	***CWI with Ensemble-Based Voting***	1	**0.8736**	**0.84**	**0.8115**
2	NILC	*Exploring Feature Engineering and Feature Learning*	3	0.8636	0.8277	0.7965
3	ITEC	*Deep Learning Architecture for CWI*	4.33	0.8643	0.8110	0.7815
4	NLP-CIC	*CWI: Convolutional Neural Network vs. Feature Engineering*	4.67	0.8551	0.8308	0.7722
5	CFILT IITB	*Towards the Effectiveness of Voting Ensemble Classifiers for CWI*	5.33	0.8478	0.8161	0.7757
6	UnibucKernel	*A kernel-based learning method for CWI*	6	0.8178	0.8127	0.7919
7	SB@GU	*SB@GU at the CWI 2018 Shared Task*	6	0.8325	0.8031	0.7832
8	TMU	*CWI Based on Frequency in a Learner Corpus*	6.33	0.8632	0.7873	0.7619
9	hu-berlin	*CWI Using Character n-grams*	9	0.8263	0.7656	0.7445
10	LaSTUS/TALN	*LaSTUS/TALN at CWI 2018 Shared Task*	10.33	0.8103	0.7491	0.7402
	Baseline	-	-	0.7579	0.7106	0.7179

ENGLISH – PROBABILISTIC TASK

Both China	and the Philippines	flexed	their	muscles	on Wednesday
0.0	0.0	0.4	0.25	0.0	0.0

	team	*paper*	avg rank	News	Wikinews	Wikipedia
1	**CAMB**	***CWI with Ensemble-Based Voting***	**1.33**	**0.0558**	**0.0674**	**0.0739**
2	ITEC	*Deep Learning Architecture for CWI*	2.33	0.0539	0.0707	0.0809
3	TMU	*CWI Based on Frequency in a Learner Corpus*	2.33	0.0510	0.0704	0.0931
4	NILC	*Exploring Feature Engineering and Feature Learning*	3.66	0.0588	0.0733	0.0819
5	SB@GU	*SB@GU at the CWI 2018 Shared Task*	5	0.1526	0.1651	0.1755
	Baseline	-	-	0.1127	0.1053	0.1112

SPANISH – BINARY & PROBABILISTIC TASKS

Jugó como	lateral		izquierdo	principalmente	con la	Juventus
simple	complex	complex	simple			complex
0.0	0.1	0.4	0.0			0.1

	team	*paper*	F_1	MAE
1	**TMU**	***CWI Based on Frequency in a Learner Corpus***	0.7699	0.0718
2	ITEC	*Deep Learning Architecture for CWI*	0.7637	0.0733
3	NLP-CIC	*CWI: Convolutional Neural Network vs. Feature Engineering*	0.7672	-
4	CoastalCPH	*Cross-lingual CWI with multitask learning*	0.7458	0.0789
5	SB@GU	*SB@GU at the CWI 2018 Shared Task*	0.7281	-
6	hu-berlin	*CWI Using Character n-grams*	0.7080	-
	Baseline	-	0.7237	0.0892

GERMAN – BINARY & PROBABILISTIC TASKS

Prellungen	und	komplizierte	Knochenbrüche	mussten	behandelt	werden
complex		complex	complex		simple	
0.333		0.167	0.25		0.0	

	team	*paper*	F_1	MAE
1	**TMU**	***CWI Based on Frequency in a Learner Corpus***	0.7451	0.0610
2	SB@GU	*SB@GU at the CWI 2018 Shared Task*	0.7427	-
3	hu-berlin	*CWI Using Character n-grams*	0.6929	-
4	CoastalCPH	*Cross-lingual CWI with multitask learning*	0.6619	0.0747
	Baseline	-	0.7546	0.0816

RESULTS

Multilingual Track

FRENCH – BINARY & PROBABILISTIC TASKS

La	construction	de l'	Atomium	fut une	prouesse		technique
	simple		complex		complex	complex	simple
	0.0		0.556		0.556	0.222	0.0

	team	*paper*	F_1	MAE
1	**CoastalCPH**	***Cross-lingual CWI with multitask learning***	**0.7595**	**0.0660**
2	TMU	*CWI Based on Frequency in a Learner Corpus*	0.7465	0.0778
3	SB@GU	*SB@GU at the CWI 2018 Shared Task*	0.6266	-
4	hu-berlin	*CWI Using Character n-grams*	0.5738	-
	Baseline	-	0.6344	0.0891

CONCLUSION

KEY TAKEAWAYS

- novel approach
  - multilingual data
  - from binary to more gradual, probabilistic CWI
- ensemble classifiers still buzzing
  - top-performing across almost all tracks
  - features similar to CWI2016 insights
- deep neural networks perform reasonably well
  - English track and need for sizable data
  - aptitude for a cross-lingual setting

THANK YOU! QUESTIONS?

www.uclouvain.be/anais.tack

anais.tack@{uclouvain,kuleuven}.be

anaïs-tack-95b869a1

REFERENCES (1/2)

AbuRa'ed, A., & Saggion, H. (2018). LaSTUS/TALN at Complex Word Identification (CWI) 2018 Shared Task. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 159–165). New Orleans, Louisiana: Association for Computational Linguistics.

Alfter, D., & Pilán, I. (2018). SB@GU at the Complex Word Identification 2018 Shared Task. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 315–321). New Orleans, Louisiana: Association for Computational Linguistics.

Aroyehun, S. T., Angel, J., Pérez Alvarez, D. A., & Gelbukh, A. (2018). Complex Word Identification: Convolutional Neural Network vs. Feature Engineering. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 322–327). New Orleans, Louisiana: Association for Computational Linguistics.

Bingel, J., & Bjerva, J. (2018). Cross-lingual complex word identification with multitask learning. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 166–174). New Orleans, Louisiana: Association for Computational Linguistics.

Brouwers, L., Bernhard, D., Ligozat, A.-L., & François, T. (2014). Syntactic Sentence Simplification for French. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* (pp. 47–56). Gothenburg, Sweden: Association for Computational Linguistics.

Butnaru, A., & Ionescu, R. T. (2018). UnibucKernel: A kernel-based learning method for complex word identification. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 175–183). New Orleans, Louisiana: Association for Computational Linguistics.

Dale, E., & Chall, J. S. (1948). A Formula for Predicting Readability. *Educational Research Bulletin*, 27(1), 11–28.

De Hertog, D., & Tack, A. (2018). Deep Learning Architecture for Complex Word Identification. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 328–334). New Orleans, Louisiana: Association for Computational Linguistics.

Elhadad, N. (2006). Comprehending technical texts: predicting and defining unfamiliar terms., Comprehending Technical Texts: Predicting and Defining Unfamiliar Terms. *AMIA Annual Symposium Proceedings*, 239–243.

Gooding, S., & Kochmar, E. (2018). CAMB at CWI Shared Task 2018: Complex Word Identification with Ensemble-Based Voting. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 184–194). New Orleans, Louisiana: Association for Computational Linguistics.

Gunning, R. (1952). *The Technique of Clear Writing*. New York: McGraw-Hill.

Kauchak, D. (2016). Pomona at SemEval-2016 Task 11: Predicting Word Complexity Based on Corpus Frequency. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 1047–1051). San Diego, California: Association for Computational Linguistics.

Hartmann, N., & dos Santos, L. B. (2018). NILC at CWI 2018: Exploring Feature Engineering and Feature Learning. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 335–340). New Orleans, Louisiana: Association for Computational Linguistics.

Kajiwara, T., & Komachi, M. (2018). Complex Word Identification Based on Frequency in a Learner Corpus. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 195–199). New Orleans, Louisiana: Association for Computational Linguistics.

REFERENCES (2/2)

Malmasi, S., Dras, M., & Zampieri, M. (2016). LTG at SemEval-2016 Task 11: Complex Word Identification with Classifier Ensembles. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 996–1000). San Diego, California: Association for Computational Linguistics.

Mc Laughlin, G. H. (1969). SMOG Grading—–a New Readability Formula. *Journal of Reading*, 12(8), 639–646.

Paetzold, G. H., & Specia, L. (2016a). Understanding the Lexical Simplification Needs of Non-Native Speakers of English. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 717–727). Osaka, Japan.

Paetzold, G., & Specia, L. (2016b). SemEval 2016 Task 11: Complex Word Identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 560–569). San Diego, California: Association for Computational Linguistics.

Popović, M. (2018). Complex Word Identification Using Character n-grams. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 341–348). New Orleans, Louisiana: Association for Computational Linguistics.

Ronzano, F., Abura'ed, A., Espinosa Anke, L., & Saggion, H. (2016). TALN at SemEval-2016 Task 11: Modelling Complex Words by Contextual, Lexical and Semantic Features. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 1011–1016). San Diego, California: Association for Computational Linguistics.

Shardlow, M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. In *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop* (pp. 103–109). Sofia, Bulgaria: Association for Computational Linguistics.

Shardlow, M. (2014). Out in the Open: Finding and Categorising Errors in the Lexical Simplification Pipeline. In *LREC 2014* (pp. 1583–1580). European Language Resources Association (ELRA).

Wani, N., Mathias, S., Gajjam, J. A., & Bhattacharyya, P. (2018). The Whole is Greater than the Sum of its Parts: Towards the Effectiveness of Voting Ensemble Classifiers for Complex Word Identification. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 200–205). New Orleans, Louisiana: Association for Computational Linguistics.

Wróbel, K. (2016). PLUJAGH at SemEval-2016 Task 11: Simple System for Complex Word Identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 953–957). San Diego, California: Association for Computational Linguistics.

Yimam, S. M., Štajner, S., Riedl, M., & Biemann, C. (2017a). CWIG3G2 - Complex Word Identification Task across Three Text Genres and Two User Groups. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 401–407). Taipei, Taiwan: Asian Federation of Natural Language Processing.

Yimam, S. M., Štajner, S., Riedl, M., & Biemann, C. (2017b). Multilingual and Cross-Lingual Complex Word Identification. In *Proceedings of Recent Advances in Natural Language Processing* (pp. 813–822). Varna, Bulgaria: Incoma Ltd. Shoumen, Bulgaria.

Zampieri, M., Malmasi, S., Paetzold, G., & Specia, L. (2017). Complex Word Identification: Challenges in Data Annotation and System Performance. In *Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2017)* (pp. 59–63). Taipei, Taiwan: Asian Federation of Natural Language Processing.